Modeling User Expertise in Folksonomies by Fusing Multi-type Features

Junjie Yao[1], Qiaosha Han[1], Bin Cui[1], Ce Zhang[2], and Yanhong Zhou[3]

[1] Department of Computer Science & Key Laboratory of High Confidence Software Technologies (Ministry of Education), Peking University
{junjie.yao, qiaoshahan, bin.cui}@pku.edu.cn
[2] Department of Computer Science, University of Wisconsin-Madison
czhang@cs.wisc.edu
[3] Yahoo! Global R&D Center (Beijing)
zhouyh@yahoo-inc.com

Abstract. The folksonomy refers to the online collaborative tagging system which offers a new open platform for content annotation with uncontrolled vocabulary. As folksonomies are gaining in popularity, the expert search and spammer detection in folksonomies attract more and more attention. However, most of previous work are limited on some folksonomy features. In this paper, we introduce a generic and flexible user expertise model for expert search and spammer detection. We first investigate a comprehensive set of expertise evidences related to users, objects and tags in folksonomies. Then we discuss the rich interactions between them and propose a unified Continuous CRF model to integrate these features and interactions. This model's applications for expert recommendation and spammer detection are also exploited. Extensive experiments are conducted on a real tagging dataset and demonstrate the model's advantages over previous methods, both in performance and coverage.

1 Introduction

Collaborative tagging is an emerging method for online content organization and management. By annotation using uncontrolled vocabulary, collaborative tagging systems provide better experience of resource sharing as well as organization. There are many sites assisted by collaborative tagging. For example, Delicious (http://delicious.com) for web page bookmarking, YouTube (http://www.youtube.com) for video sharing, Flickr (http://www.flickr.com) for photo sharing, and Twitter (http://www.twitter.com) with hashtag. This collaborative organization approach is also called *folksonomy*.

Along with the developments of these tagging systems, many research problems have been studied to improve folksonomy. For example, personalized recommendation is discussed in [4], an improved tag based content retrieval is presented in [13], and one of our previous work in [2] exploits novel features fusion methods for tagged resources.

As tagging systems gain in popularity, experts and spammers flow into the tagging sites at the same time. User interaction becomes difficult, and finding appropriate information is urgent. In this paper, we study the problem of modeling users' expertise in collaborative tagging communities. That is to discover the user's expertise with respect to a given topic (or a tag), of which can be made use to clearly distinguish the experts and spammers.

The expert search problem, which has already caught our eyes in enterprise corpora [1] and recently social networks [6], is also very meaningful to current tagging communities. Nevertheless, without the ability to combat spammers, the system will suffer the misleading influences of them. For example, an expertise model is helpful in the case that a user wants to find top experts on a specific topic and then follow their activities. With a suitable expertise model, we can also directly recommend a user the experts of certain topics which may be interesting to him/her. However, the expert list recommended by the system for the user will be filled with some useless spammers if we cannot eliminate them accurately. What is more, a suitable expertise model can also improve tag qualities. On one side, tagging systems can directly push the resources with imprecise tags to the expert users and let them tag. On the other side, we can avoid the misleading influences of spammers on tag quality calculation, by simultaneously distinguishing spammers from users. We believe that, with the help of a reliable expertise model to search experts and combat spammers, a lot of applications as those we mention above can be supported in such communities to improve user satisfaction significantly.

The most crucial difference between folksonomy and traditional classic enterprise corpus or social network is that the former has a comprehensive set of features, e.g., users, tags and objects. There also exist rich and meaningful interactions among them. For example, it is not surprising to see that the expertise of a certain user on tag t is determined not only by his/her tagging behavior on t, but also on tags similar to t, or the user's social network involved. Such multi-type features can not only help us get more reasonable expert ranking, but also better protect our system from malevolent attacks of spammers.

To the best of our knowledge, there do not exist many solutions to deal with the expertise modeling problem in folksonomies. Current methods usually utilize a part of information about users, objects and tags [15, 5, 11]. However, no work have explored all features related to users, objects and tags in folksonomies. We believe this could result in better representation models, and further make folksonomies more accurate in expert finding and more resistant to spammers.

To fully utilize existing multi-type features, we propose a novel expertise model for collaborative tagging communities. We extract several expertise evidences/features hidden in folksonomies. An expertise model is used to combine those expertise evidences and generate users' expertise over topics/tags. Experiments demonstrate the advantage of this integrated model.

We outline the contributions of this paper as below:

1. We extract a comprehensive set of expertise evidences/features hidden in such tagging communities. Considering the fact that collaborative tagging communities basically consist of three parts: tags, users, and objects, those evidences can be classified into three categories similarly: 1) tag-related evidences, 2) user-related evidences and 3) object-related evidences.
2. The expertise model based on Continuous Conditional Random Fields (CRF) [12] is introduced to automatically integrate those expertise evidences and generate

user's expertise over topics/tags as a result. This model is inspired by the successful applications of CRF technique [8] to model interactions between different items in an undirected graph. As for our expertise modeling problem, the proposed model can keep the balance among different kind of features and also make full use of them.
3. Our experiments conducted on the expertise ranking problem in a real tagging dataset show clearly that our CRF-based expertise model is obtaining much higher precision on searching experts and more resistant to spamming activities than any other baselines.

The rest of this paper proceeds as follows. In Section 2 we formulate the problem and explore the evidences inspiring our expertise model. Section 3 presents the CRF-based expertise model. We discuss the modeling and learning of this model in details. Quantitative experiments are shown in Section 4. After Section 5 reviews the related work, we conclude this paper in Section 6.

2 Expertise Evidence in Folksonomy

In this section, we carry out a thorough analysis of the folksonomies and investigate all evidences which are helpful to our expertise model. We begin with the problem formulation of expertise modeling task and correlated structure of features in folksonomy. Then a comprehensive set of expertise evidences are studied.

2.1 Feature Correlation

Let $X = \{O, U, T, R\}$ denote all the observations we have in a folksonomy including all the objects $O = \{o_1, o_2, \ldots, o_M\}$, users $U = \{u_1, u_2, \ldots, u_N\}$, tags $T = \{t_1, t_2, \ldots, t_L\}$ and their relationships $R = \{(o_i, u_j, t_k)\}$. Also, define a matrix $\mathbf{E} = \{e_{ij}\}$, where e_{ij} denotes the expertise score of user u_i on tag t_j. With higher value of e_{ij}, user u_i is more likely to be an expert on tag t_j.

The expertise modeling problem is to determine the expertise score matrix $\mathbf{E}$ given all the observations X in a folksonomy. We need to infer reasonable expertise scores based on all observations. Hence, our task can be modeled to find $\mathbf{E}$ that maximizes the appearance probability of $\mathbf{E}$ given current observations X, i.e.,

$$\mathbf{E} = \arg\max_{\mathbf{E}} P_\theta(\mathbf{E}|X) = \arg\max_{\mathbf{E}} P_\theta(\mathbf{E}|O, U, T, R), \quad (1)$$

where θ denotes the model parameter. However, the computing $P_\theta(\mathbf{E}|O, U, T, R)$ is not a trivial task, as discussed in the above section.

To model the interactions among users, objects and tags and show the influences of the interactions on expertise, we introduce a graph structure, in Figure 1. Shown in this graph, three core elements: users, objects and tags are represented by grey nodes and expertise scores by blank nodes in the upper part. Here we present relations among the same type of nodes as well as the cross-type relations. For example, the edge between user u_i and u_j stands for a subscribe/as-a-friend relation in the tagging system. And for

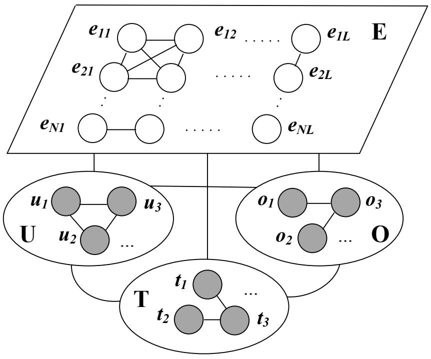

Fig. 1. Graph Structure of Folksonomy

the cross-type relation, the use of tag t_j by user u_i is represented by an edge between the corresponding nodes of t_j and u_i. Edges in this graph are weighted. In this scenario, the more frequently user u_i uses tag t_j, the larger the edge weight between node u_i and node t_j is.

Besides those observations $X = \{O, U, T, R\}$, our expertise scores, which are denoted in the top layer **E**, are the output of our expertise model. Each node e_{ij} in this layer represents the expertise of user u_i on tag t_j. In this graph the expertise nodes are interlinked. This aligns with our intuition that there are hidden interactions among the expertise scores of different users on different tags.

To present these mutual influences, each node in the expertise layer is connected to the nodes corresponding to observations in folksonomy, which implies that a user's expertise in folksonomy is influenced by all items, i.e., users, objects and tags. The intuitive meanings of these edges will be clear after explaining the *Expertise Evidences* we find.

2.2 Expertise Evidences

Only the structure of folksonomy can not give us enough information about the user expertise, here we illustrate more facets and evidences in folksonomy for more inspirations to estimate expertise reasonably.

The *Expertise Evidences* refer to the information which can help discover experts and tell spammers in a folksonomy system. Some state-of-the-art work have already proposed some indications, e.g., user authority [15], tag reliability [5] and post date [11]. However, they usually fail to conduct a comprehensive study of feature interactions, which are actually the most crucial difference between folksonomy and classic web environment or the enterprise corpus. In contrast, we first investigate an overall set of *Expertise Evidences*. Considering the fact that users, objects, and tags are three basic foundations of folksonomy, we observe the system from these three perspectives and explore the corresponding evidences.

We categorize our discovered evidences into three types: 1) tag-related evidences, 2) object-related evidences and 3) user-related evidences.

Tag-related Evidences Two tag-related expertise evidences are applied in our model.

The first evidence suggests that, if a user often agrees with others on the choice of tag to label on an object, there is a great chance that he/she is an expert. This is consistent with our intuition [15, 5], however, this feature alone is very vulnerable to attacks conducted by spammers as shown in our following experiments. *(TE-1)*

The second one is that an expert on tag t should have high expertise on similar tag t', too. Imagine an expert on topic *Web2.0*, even though he/she may not tag many tags on similar topics, e.g., *Delicious*, we can trust these tags because that his/her knowledge has been reflected on a very similar topic. This evidence may also help us handle the different personal customs on tagging. For example, tell an expert on *Puppy* in spite of his preference to use *Dog* when tagging. *(TE-2)*

User-related Evidences The user-related evidence comes from the subscribe relation in folksonomy. Two metrics are used to depict user's characteristics: the expertise on certain tag and the ability to find an expert on certain tag.

In the directed subscribe relation, the difference and interaction between the two metrics are important. The expert user tends to be subscribed by users who are good at discovering experts, and the user with high ability of discovering experts is more likely to subscribe many experts. This is reasonable for us, given that in practice experts are often subscribed by others because they often provide useful knowledge. Also a user subscribing lots of experts is good at finding experts because his behavior is a good sign of the ability to distinguish the usefulness of provided knowledge. These two sub-rules together show the mechanism of the interactive influence between user's expertise and user's ability of finding experts. *(UE-1)*

Object-related Evidences It is always not easy to analyze the content of objects, and hence we will not use object's content information in this work. Instead, we should notice that the user's expertise on tag t should be increased if he/she uses tag t to annotate the object o which later becomes a popular object.

This evidence combines the popularity of an object and the post time information. It satisfies the assumption in [11] that an expert should have the ability of discovering popular objects instead of just following others. However, compared to [11], we model this feature in a different way to integrate it into our model seamlessly. *(OE-1)*

For any folksonomy system, the three elements, i.e., users, objects and tags, are pillars of the pyramid, and all the facts and evidences we are considering have fully covered the three main foundations of the folksonomy. The evidences and rules directly resulting from these facts are laconic, easy to obtain, reasonable intuitively at the same time. It cannot be denied that other features can also be drawn from the folksonomy graph structure we presented above, too. For example, one can write *ad hoc* evidences specific to application scenario or user preference. But we need to point out that the four basic evidences we pick now can fully cover all three facets of a folksonomy system and as shown in our latter experiment, our model integrating barely these four evidences can still beat other state-of-the-art techniques and perform satisfactorily. What is more important, any other evidence can be easily integrated into our model. As a result, in this paper, we can mainly focus on these evidences and show how they can be integrated into a unified model.

3 CRF based Expertise Model in Folksonomy

In this section, we present our CRF based expertise model in detail. The CRF based expertise model is proposed to fuse multi-type features in folksonomy and to finally generate expertise scores of users on specific tags. We formulate the relations in this unified model and also discuss the parameter learning methods.

3.1 Model Formulation

In order to well fuse the expertise evidences discussed in Section 2, a Continuous Conditional Random Fields [12] based expertise model is proposed here to model the user expertise in folksonomy. Compared to other fusion models or heuristic methods, this model is powerful in automatic feature weighting and interaction combination. To cope with requirements in this problem setting, we also discuss the improvement over basic CRF model.

Recall the problem defined above, we aim at estimating the probability of $\mathbf{E}$ given the observation $X = \{O, U, T, R\}$, and hence we can maximize this probability to obtain an optimal $\mathbf{E} = \arg\max_{\mathbf{E}} \{P_\theta(\mathbf{E}|X)\}$. Continuous CRF provides a way to estimate such probability:

$$P_\theta(\mathbf{E}|X) = \frac{1}{Z(X)} \exp\left\{\sum_{i=1}^{k} \left(\boldsymbol{\lambda_o} \cdot \boldsymbol{F_o}(c_i) + \boldsymbol{\lambda_u} \cdot \boldsymbol{F_u}(c_i) + \boldsymbol{\lambda_t} \cdot \boldsymbol{F_t}(c_i)\right)\right\}, \quad (2)$$

where $\{c_i | i \in [1, k]\}$ is a set of k cliques in our graph. For example, in Figure 1, $c = \{u_1, t_1, t_2, e_{11}, e_{12}\}$ can be taken as a clique if the corresponding nodes of these elements in the set are fully connected. In the above equation, $\boldsymbol{F_o}$ denotes a function vector consisting of *feature functions* designed for the object-related evidences and $\boldsymbol{\lambda_o}$ represents a weight vector of those features or evidences. $\boldsymbol{F_u}$, $\boldsymbol{\lambda_u}$, $\boldsymbol{F_t}$ and $\boldsymbol{\lambda_t}$ all have similar definitions. The variable θ stands for the parameter set of this model: $\theta = \{\boldsymbol{\lambda_o}, \boldsymbol{\lambda_u}, \boldsymbol{\lambda_t}\}$ which satisfies $\sum_i \lambda_o^i + \sum_j \lambda_u^j + \sum_k \lambda_t^k = 1$ and $Z(X)$ is a normalization factor defined as

$$Z(X) = \int_{\mathbf{E}} \exp\left\{\sum_{i=1}^{k} \left(\boldsymbol{\lambda_o} \cdot \boldsymbol{F_o}(c_i) + \boldsymbol{\lambda_u} \cdot \boldsymbol{F_u}(c_i) + \boldsymbol{\lambda_t} \cdot \boldsymbol{F_t}(c_i)\right)\right\}. \quad (3)$$

We need to define different *feature functions* for the three kinds of evidences to integrate them into this model. Here the *feature function* refers to a function defined on clique c to measure the fitness of nodes in c to appear together. Specifically, four feature functions are designed with respect to the four expertise evidences stated above. Before we explain the detailed *feature functions*, we first define variables extracted to describe information in folksonomy.

- **Tag similarity matrix** $\mathbf{S}_{tag}$: each entry $\mathbf{S}_{tag}(t_i, t_j)$ equals the similarity between tag t_i and t_j, which is calculated by tag co-occurrence in our implementation.
- **User subscription matrix** $\mathbf{Sub}$: $\mathbf{Sub}(u_i, u_j) = 1$ iff. user u_j subscribes user u_i.
- **User temporal similarity matrix** $\mathbf{S}_T$: $\mathbf{S}_T(u_i, u_j|t_k)$ is the similarity between u_i and u_j tagging behaviors computed based on the average number of users who follow u_i and u_j on objects tagged by t_k.
- **Expertise matrix E**: each entry e_{ij} represents the expertise score of user u_i on tag t_j.
- **Expert finding ability matrix** $\mathbf{E}'$: each entry e'_{ij} denotes the ability of user u_i to find experts on tag t_j.

We then write feature functions according to the suggested evidences above. Note that, these feature functions only have non-zero values to certain types of cliques and are automatically set to zero for cliques of other kinds.

– **TE-1:** We define f_t^1 on clique c like $\{e_{ij}, u_i, t_j\}$ as

$$f_t^1(e_{ij}, u_i, t_j) = -\Big(e_{ij} - \mathcal{N}\big(\sum_{o_{t_j}} CoIn(u_i, t_j)\big)\Big)^2, \tag{4}$$

where o_{t_j} enumerates the objects tagged by u_i with tag t_j, and $CoIn(u_i, t_j)$ is the number of users who agree with user u_i to apply tag t_j to object o_{t_j} for tag t_j. Function $\mathcal{N}(.)$ is introduced to normalize the input variable to $[0, 1]$. This feature represents the evidence that if one user agrees with more other users on a certain tag, his/her expertise on this tag should be higher.

– **TE-2:** We define f_t^2 on clique c like $\{e_{ij}, e_{ik}, t_j, t_k\}$ as

$$f_t^2(e_{ij}, e_{ik}, t_j, t_k) = -\frac{1}{2(|T| - 1)}\mathbf{S}_{tag}(t_j, t_k) \times (e_{ij} - e_{ik})^2, \tag{5}$$

where $|T|$ is the number of all tags in the system. By this feature function, the user u_i's expertise scores on similar tags t_j and t_k would be close.

– **UE-1:** We define f_u^1 on clique c like $\{e_{ij}, e'_{kj}, u_i, u_k\}$ as

$$f_u^1(e_{ij}, e'_{kj}, u_i, u_k) = -\mathbf{Sub}(u_i, u_k) \times (e_{ij} - e'_{kj})^2, \tag{6}$$

where $e'_{kj} = \sum_{i=1}^{|U|} \big(\mathbf{Sub}(u_i, u_k) \times e_{ij}\big)$ and $|U|$ is the number of all users in folksonomy. This user-related feature function encodes the two-side interactions between expertise score and finding expert ability into a unified framework.

– **OE-1:** We define f_o^1 on clique c like $\{e_{ij}, e_{kj}, u_i, u_k, t_j\}$ as

$$f_o^1(e_{ij}, e_{kj}, u_i, u_k, t_j) = -\frac{1}{2}\mathbf{S}_T(u_i, u_k|t_j) \times (e_{ij} - e_{kj})^2. \tag{7}$$

This means it is better for two users u_i and u_k to own similar high expertise on tag t_j if they both label a popular object with tag t_j and they discover the object earlier than most other users.

3.2 Parameter Learning

The learning process of our expertise model is to obtain parameters $\theta = \{\boldsymbol{\lambda_o}, \boldsymbol{\lambda_u}, \boldsymbol{\lambda_t}\}$, given a training dataset $D = (X, \mathbf{E})$, X includes objects O, users U, tags T and their relations R. In matrix $\mathbf{E} = \{e_{ij}\}$, each entry e_{ij} represents the expertise score of user u_i on tag t_j. We normalize the expertise scores to $[0, 1]$.

One traditional technique for parameter learning is to train a model which can maximize log-likelihood of training dataset D's appearance. There exist lots of discussions about how to learn the optimal parameters in CRF framework, e.g., *Gibbs Sampling* from [14]. However, it may not optimize the desired objective function, i.e., the average precision of expertise ranking problem in our case. In contrast, direct optimization aiming at the evaluation metric is better in some scenarios [10]. In this paper, we use the methodology applied in [10]. Specifically speaking, we enumerate the combination of parameter θ and select parameter which makes the model obtain the maximal average precision of expert ranking task.

In our problem setting, we are only interested in ranking users by their expertise, so the inference process can be simplified. The $Z(X)$ will influence only the absolute expertise scores, not the ranking positions. Under this occasion, the ranking score matrix of users on certain tags is denoted by $\mathbf{E}'$.

$$
\begin{aligned}
\mathbf{E}' &\propto \arg\max_{\mathbf{E}} \exp\left\{\sum_{i=1}^{k} \big(\boldsymbol{\lambda_o} \cdot \boldsymbol{F_o}(c_i) + \boldsymbol{\lambda_u} \cdot \boldsymbol{F_u}(c_i) + \boldsymbol{\lambda_t} \cdot \boldsymbol{F_t}(c_i)\big)\right\} \\
&\propto \arg\max_{\mathbf{E}} \sum_{i=1}^{k} \big(\boldsymbol{\lambda_o} \cdot \boldsymbol{F_o}(c_i) + \boldsymbol{\lambda_u} \cdot \boldsymbol{F_u}(c_i) + \boldsymbol{\lambda_t} \cdot \boldsymbol{F_t}(c_i)\big)
\end{aligned}
$$

After substituting the detailed feature functions, i.e., tag related functions: $f_t^1(.)$, $f_t^2(.)$, user related function: $f_u^1(.)$, as well as object related function: $f_o^1(.)$, into this equation, we can generate the solution of $\mathbf{E}'$ by using standard Lagrange multiplier methods.

4 Empirical Study

In this section, we evaluate our expertise model by the expertise ranking problem in folksonomies. Specifically, we conduct the evaluations on *expert ranking* and *spammer ranking* to answer the following two questions respectively:

Q1: How exactly is the performance of our expertise model on searching experts for specific tags?

Q2: Is the proposed expertise model robust enough to resist the spammers' attacks?

4.1 Experimental Setup

Experiments are conducted on a real tagging dataset, collected from Delicious (http://delicious.com/) website. These tags range from Jan. to Jun. 2010. The dataset contains

$10,800,690$ web page urls, $197,783$ users, $1,928,677$ tags. We also fetch subscription relations between users.

The distribution of tag frequency is shown in Table 1. It is easy to tell that, less tags show in the dataset with higher frequency. In our experiments, we mainly focus on the tags ranging from level-3 to level-5. In truth, more people will be interested in experts on tags such as "asp.net" in level-4 than those like "vibes" in level-1. Since users' interests mainly focus on popular tags, these tags deserve more attention. And the improvement in experts search on these tags can dramatically enhance user satisfaction.

Table 1. Statistics of tag frequency

Level-ID	Frequency Interval	#Tags
0	[1,9]	1,694,768
1	[10, 99]	199,563
2	[100,999]	28,557
3	[1000,9999]	4,921
4	[10000,99999]	780
5	[100000,999999]	88

Training Set and Testing Set To construct our testing query set, we randomly select 10 tags for each frequency level from level-3 to level-5. This is the base to run the model and to evaluate its performance on *expert ranking* and *spammer ranking*.

For *expert ranking* part, parameter learning is crucial to our expertise model as illustrated before. To learn the parameter set $\theta = \{\boldsymbol{\lambda}_o, \boldsymbol{\lambda}_u, \boldsymbol{\lambda}_t\}$, a small training set is manually annotated by two annotators. Annotators are asked to assign the binary expertise score (expert or not) to randomly selected users to some randomly selected tags. With the annotation result as ground truth, we adjust the parameter to achieve higher expert recommendation precision.

Turning to *spammer ranking* part, in order to measure how resistant the expertise model is to the spammers' attacks, we follow the method used in [11]. Three types of spammers are randomly inserted, i.e., *Flooders*, *Promoters* and *Trojans*. *Flooders* refer to users who tag a extremely large number of tags, while *Promoters* always tag their own web pages and pay little attention to objects provided by other users. Much more crafty, *Trojans* tag a lot for their own pages but at the same time conceal their malicious intentions by acting like regular users. More details can be found in [11]. Specifically, each query tag has 20 spammers of each type.

Evaluation Metrics We apply the Precision@N as our main evaluation metric, which represents the percentage of answers that are "correct" in all N candidate answers retrieved, taking the manually annotated experts list and inserted spammers list as ground truth.

In particular, in *expert ranking* part, the retrieved user is "correct" if he/she is labeled as expert by annotators. As for *spammer ranking* task, the retrieved user is "correct" if he/she is a simulated spammer we inserted. Hence, the higher Precision@N for expert ranking, the more reliable and suitable the model. In contrast, a higher Precision@N for spammer ranking task is an indication that the model is more vulnerable to spammers' attack activities.

In addition to the Precision@N, we use another metric, i.e., Average Ranking Position, to measure the difference in model's ability to demote spammers in the expert list by giving spammers lower scores than true experts. The higher the metric, the more resistant the expertise model to spammers' attacks.

Baseline Methods We compare our method with three state-of-the-art approaches for both expert ranking and spammer ranking.

- Baseline 1: ***HITS*** [15]. It applies HITS algorithm to determine the user expertise by assuming that there exit reciprocal reinforcements between user expertise and tag quality.
- Baseline 2: ***CoIn*** [5]. It uses the coincidence between users as the expertise of users.
- Baseline 3: ***SPEAR*** [11]. It assumes the users tagging objects of more popularity or tagging objects earlier deserve higher expertise scores.

4.2 Quantitative Result

We report our performance study on two tasks, i.e., *expert ranking* and *spammer ranking*.

Expert Ranking We present our experimental results to answer how exactly the performance of our expertise model on searching experts is. Table 2 shows the results of different approaches, including our method named Multi-type Feature Fusion (MFF) and three baselines, i.e., SPEAR, CoIn and HITS. We have tried different popular levels of tags, however they obtain similar results so we do not report them separately.

As seen in this table, in respect of Precision@1, our MMF is tied with other baselines. Also, our MFF approach obtains the best performance in top-5 precision, however, it meets a slightly decrease in performance for larger N. From the overall perspective HITS obtains the best performance but our approach is the second best.

Several facts can interpret this result. First of all, due to the limited time, when the annotators determine whether the retrieved user is an expert on the query tag, they usually take the tag frequency of users as the most important factor, but true users in a collaborative tagging system will consider more in fact. Hence, the annotated results may be more inclined to HITS method. Secondly, as reported by the annotators, they can not easily determine whether some users with tremendous tags are spammers or not. Generally, they just take those users as experts instead of spammers.

Despite having slightly worse overall performance than HITS when measured by annotated results and assuming all manually annotated experts are accurate, our model can improve user satisfaction empirically and also, the top 5 retrieved users of our

Table 2. Average Precision for Expert Ranking Task

Average Precision	SPEAR	CoIn	HITS	MFF
$P@1$	**1**	**1**	**1**	**1**
$P@5$	0.80	0.96	0.96	**1**
$P@10$	0.68	0.90	**0.92**	0.88
$P@15$	0.64	0.85	**0.91**	0.85
$P@20$	0.61	0.87	**0.91**	0.88

method for all query tags are all annotated as experts. We believe in fact, our algorithm can averagely achieve satisfactory results, and can be better or at least comparable to most state-of-the-art approaches.

Spammer Ranking This experiment is conducted to test how our expertise model's performance is when confronting malicious spammers.

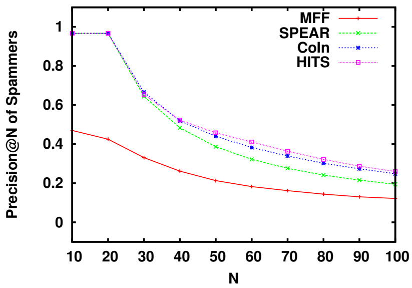

Fig. 2. Average Precision of Spammer Ranking, from level-3 to level-5

When measuring with the metric of Precision@N, Figure 2, 3 and 4 present the performance of *spammer ranking* for different models here. The overall performance of whole query set is shown in Figure 2, while Figure 3 and 4 give more detailed information about the model's performance concerning the discrepancies of frequency levels in tags. The result for the level-4 query tags is quite similar to level-5, so we do not show it.

In addition, we show the evaluation results in Figure 5 for different expertise models with respect to various types of spammers. The y-axis represents the average ranking position of specified type spammers in top $10,000$ retrieved users. In Figure 2 and Figure 3, for the *spammer ranking* task, the lower value in Precision@N suggests there are less spammers in the first N experts, which means the expertise model is more resistant to spammers' attacks. In contrast, in Figure 5, the lower value in y-axis serves as an indication that inserted spammers are decided to be experts by the model with high expertise scores, showing the system's vulnerability to spammers' attacks.

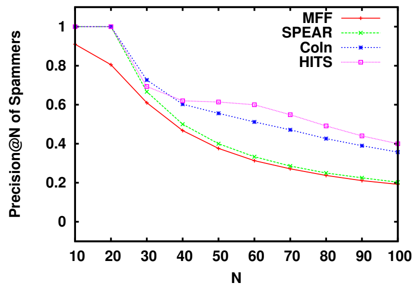

Fig. 3. Average Precision of Spammer Ranking, level-3

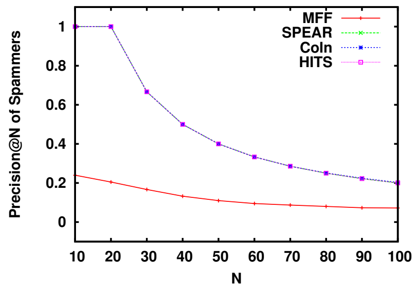

Fig. 4. Average Precision of Spammer Ranking, level-5

In Figure 2, the average Precision@20 almost equals 1 for three baselines, but only approximately 0.4 for our approach. With the increase of the recommended expert number, the advantage of our model shrinks. However, even at Precision@100, our approach is still better than the other three. As discussed later, it is the seamless fuse of different types of expertise information that makes our method most resistant to spammers on average.

According to Figure 3 and Figure 4, we can find that all the expertise models can obtain better performance when dealing with tags of higher popularity. One possible reason might be that for the tags with less popularity, the spammers are easier to "beat" regular users and become the "experts" on certain topics. Figure 5 suggests that no

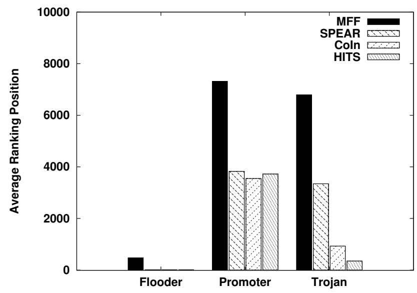

Fig. 5. Average Spammer Ranking Positions

matter what kind the spammers belong to, our expertise model MMF shows great improvement in the ability to demote all the spammers from the top of the expert list. To be more specific, all three baselines show nearly no resistance to *Flooders* while MMF goes a big step further than them. As for *Promoters* and *Trojans*, these spammers will be demoted twice or more in the expert list by MMF than by any other method in the three.

Result Analysis and Discussion From the above expert ranking and spammer ranking experiments, we can see that our MFF approach outperforms the baseline models. And we should point out that the good performance of MMF approach results from its ability to integrate multi-type information extracted from the folksonomy system.

When distinguishing experts, we have more clues and features to inspire experts search. So we can get accurate experts list especially when most top experts are needed. While telling spammers, CoIn and HITS both cannot well separate *Trojans* from regular users. This failure is a result of their too much dependence on the tag frequency information. Although when added to the temporal information, SPEAR has great improvements in telling *Trojans* and basically performs better than CoIn and HITS, it still cannot identify the *Flooders* like MMF because *Flooders* tag too many resources to be distinguished by temporal information easily. However, our MMF model integrates information of a wider range, e.g., subscription network among users, yielding more satisfactory results both on the average spammers ranking and spammers demotion of special kinds.

Considering the performance of our expertise model on *expert ranking* and *spammer ranking*, the model is believed to suitable for practical applications in real world collaborative tagging systems. First, although the model is not best when recommending large scale of experts, but in our daily life, only the top experts recommended are interesting to users. Too much patience are needed for a user to browse 10 or more recommended experts everyday. Second, the outstanding performance on *spammer ranking* means our recommended expert list will not be filled with spammers, especially *Flooders*. Hence, compared to other methods, such as HITS, our expertise model is good at demoting crafty spammers like *Flooders*, who will waste user's energy to follow and reduce user satisfaction significantly. Given all these reasons, by accurate top experts ranking without misleading spammers in the top positions, our expertise model will provide satisfactory services in real world folksonomies.

5 Related Work

Our work in this paper is broadly related to several areas. We review some in this section.

Social Media Management: With the recent startling increase of social media application, the uncontrolled vocabulary annotation is becoming popular. Researchers have discussed various directions of tagging systems. A structured tag recommendation approached was discussed in [4], and [13] presented an improved retrieval algorithm based on sequential tags. One of our previous work [2] proposed a feature fusion approach for social media retrieval and recommendation. Work in this paper focuses on a unified model for both spammer detection and expert recommendation.

Spammer Detection: Another line of related work is spammer detection which aims to detect the spammers in collaborative tagging systems or other similar systems. Here we do not pay much attention to explicitly illustrate the details in spammer detection work, instead, we focus more on the information utilized in these work. In [9], the classification methods were utilized to differentiate spammers from regular users. The features used in those machine learning method mainly focused on the content of resources. Another example was [7], in which co-occurrence information of tags and resources was used to detect spammers. In detail, the manually annotated spammer scores were propagated through a user graph, the edges of which were generated from the co-tagging, co-resource and co-tag-resource relations among different users. However, the system suffered from the problem of human labeled training data, which limited the use of the system in large scale data. Also, in [15], Xu et al. applied HITS algorithm on the bipartite graph of users and tags to implement the mutually reinforcement between tag qualities and user authorities. In addition, when dealing with tag recommendation problem in [5], the authors measured user's reliability by the frequency of the user's tags agreeing with other users' postings.

Expert Recommendation: With the widespread use of social communities in our daily life, online user modeling and expert recommendation or expert search show its importance. Researchers have made great efforts towards this direction. For example, [16] explored expertise networks in online systems, user interest and expertise modeling in social search engine was discussed in [6]. Usually, content based and structure based methods are used in user profiling. Expert search task in enterprise corpora is always of interest for many researchers. There exist two seminar models applied, i.e., document based model [3] and profile based model [1]. In one of our previous work, we combine the profile and structure based method together for community expert recommendation [17].

To tell experts in folksonomies, Noll et al. focused on structure property and proposed a HITS based algorithm on the bipartite graph among users and objects graph to extract users' expertise information in [11].

Different from all existing methods for expert ranking in tagging systems, we introduce a new expertise model to integrate a comprehensive set of expertise evidences among users, tags, and objects. With this fusion framework, our method can obtain better performance than those state-of-the-art approaches, both in combating spammer and expert ranking.

6 Conclusion

In this paper, we have addressed the problem of modeling users' expertise in folksonomies by fusing multi-type features. Compared to state-of-the-art methods, we highlighted coding the multiple interactive evidences into a unified framework by employing Continuous Conditional Random Fields techniques.

We examined the performance of our method in large scale real-world tagging data both about expert ranking and spammer ranking. According to our experiments, we find our proposed model obtains high precision in expert ranking problem in folksonomies and is also far more resistant to the spamming attacks than those state-of-the-art approaches.

We plan to extend our expertise model in two aspects. First, we will further investigate more evidences from real world folksonomies while considering the characteristics of different social sites. Second, we will employ our expertise model to facilitate other applications in folksonomies, e.g., tag-based retrieval or tag ranking.

Acknowledgements This research was supported by the National Natural Science foundation of China under Grant No. 61073019, 60933004 and 60811120098.

References

1. K. Balog, L. Azzopardi, and M. de Rijke. A language modeling framework for expert finding. *Information Processing and Management*, 45(1):1–19, 2009.
2. B. Cui, A. Tung, C. Zhang, and Z. Zhao. Multiple feature fusion for social media applications. In *Proc. of ACM SIGMOD*, pages 435–446, 2010.
3. H. Deng, I. King, and M.R. Lyu. Enhancing expertise retrieval using community-aware strategies. In *Proc. of ACM CIKM*, pages 1733–1736, 2009.
4. Z. Guan, J. Bu, Q. Mei, C. Chen, and C. Wang. Personalized tag recommendation using graph-based ranking on multi-type interrelated objects. In *Proc. of ACM SIGIR*, pages 540–547, 2009.
5. P. Heymann, G. Koutrika, and H. Garcia-Molina. Fighting spam on social web sites: A survey of approaches and future challenges. *IEEE Internet Computing*, 11(6):36–45, November 2007.
6. D. Horowitz and S. Kamvar. The anatomy of a large-scale social search engine. In *Proc. of WWW*, pages 431–440, 2010.
7. R. Krestel and L. Chen. Using co-occurence of tags and resources to identify spammers. In *Proc. of ECML PKDD Discovery Challenge Workshop*, pages 38–46, 2008.
8. J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In *Proc. of ICML*, pages 282–289, 2001.
9. A. Madkour, T. Hefni, A. Hefny, and K.S. Refaat. Using semantic features to detect spamming in social bookmarking systems. In *Proc. of ECML PKDD Discovery Challenge Workshop.*, pages 55–62, 2008.
10. D. Metzler and W.B. Croft. A markov random field model for term dependencies. In *Proc. of ACM SIGIR*, pages 472–479, 2005.
11. M.G. Noll, A. Yeung, et al. Telling experts from spammers: expertise ranking in folksonomies. In *Proc. of ACM SIGIR*, pages 612–619, 2009.
12. T. Qin, T. Liu, X. Zhang, D. Wang, and H. Li. Global ranking using continuous conditional random fields. In *Proc. of NIPS*, pages 1281–1288, 2008.
13. N. Sarkas, G. Das, and N. Koudas. Improved Search for Socially Annotated Data. *PVLDB*, 2(1):778–789, 2009.
14. X. Xin, I. King, H. Deng, and M.R. Lyu. A social recommendation framework based on multi-scale continuous conditional random fields. In *Proc. of ACM CIKM*, pages 1247–1256, 2009.
15. Z. Xu, Y. Fu, J. Mao, and D. Su. Towards the semantic web: collaborative tag suggestions. In *Proc. of WWW Collaborative Web Tagging Workshop*, 2006.
16. J. Zhang, M.S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *Proc. of WWW*, pages 221–230, 2007.
17. Y.H. Zhou, G. Cong, B. Cui, C.S. Jensen, and J.J. Yao. Routing Questions to the Right Users in Online Communities. In *Proc. of ICDE*, pages 700–711, 2009.